Alphabetized co-authorship in economics reconsidered

Klaus Wohlrabe and Lutz Bornmann

10 May 2019

Alphabetized co-authorship in economics reconsidered

Klaus Wohlrabe*
ifo Institute – Leibniz-Institute
for Economic Research
at the University of Munich e.V.
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1229
wohlrabe@ifo.de

Lutz Bornmann
Division for Science
and Innovation Studies
Administrative Headquarters
of the Max Planck Society
Munich, Germany

* Corresponding author.

Abstract

In this article, we revisit the analysis of Laband and Tollison (2006) who documented that articles with two authors in alphabetical order are cited much more often than non-alphabetized papers with two authors in the *American Economic Review* and the *American Journal of Agricultural Economics*. Using more than 120,000 multi-authored articles from the Web of Science economics subject category, we demonstrate first that the alphabetization rate in economics has declined somewhat over the last decade. Second, we find no statistically significant relationship between alphabetized co-authorship and citations in economics (the coefficients are very small). Third, we show that the likelihood of non-alphabetized co-authorship increases the more authors an article has.

JEL Code: A12, A14
Keywords: alphabetization, co-authorship, citations, Web of Science

1 Introduction

Citations are frequently used to evaluate the usefulness of research. The novelty and popularity of research, for example, is reflected in how certain ideas are taken up by colleagues and become part of the established knowledge in a field. On the other hand, citations are key factors for measuring personal and institutional success in academia. In particular, citation counts, journal impact factors (JIFs, Clarivate Analytics) and journal rankings (Bornmann, Butz, and Wohlrabe 2018) are key elements for authors' and institutions' rankings. They have an immediate impact on the career of economists as a means of competition on the academic job market. Further, they serve as an important way of indicating the quality of research when it comes to decisions concerning tenures or the appropriation of research grants. However, factors other than quality potentially affect the number of citations a particular document receives (Tahamtan and Bornmann 2018).

Laband and Tollison (2006) (henceforth LT) investigated one of these potential effects, that of alphabetized co-authorship on citations. LT use a simple ordinary least squares (OLS) regression approach by controlling several aspects that might affect the citation rate of an article, i.e. number of pages, authors, tables and self-citations. LT showed that articles with two authors in alphabetical order accrue significantly more citations than those which are not. However, this finding does not hold true for three or more authors of an article. They conclude that the optimal team size for articles is two. LT based their investigation on a small sample consisting of the *American Economic Review* and the *American Journal of Agricultural Economics*, using data from the 1980s and 1990s, respectively. But does their conclusion hold true for other journals, e.g. less prestigious journals, and more recent articles? Is there a general relationship between author ordering and citations in economics?

In this article, we revisit and extend LT's analysis using a much larger data set of more than 120,000 multi-authored articles from 1990 to 2013 published in 307 economics journals. The article is structured in two parts. First, we offer an overview of alphabetization patterns across both time and journals. Second, we investigate whether there is a relationship between alphabetized co-authorship and citations. We employ four different regression settings to

ensure the robustness of the results. Third, we investigate the relationship between alphabetized co-authorship and citations for each journal in our sample and over time. The article is structured in two parts. First, we offer an overview of alphabetization patterns across both time and journals. We then investigate whether there is a relationship between alphabetized co-authorship and citations. We employ four different regression settings to ensure the robustness of the results. Additionally, we investigate the relationship for each journal in our sample and over time. Finally, we discuss our results.

2 Data and descriptive statistics

In reflection of the core results by LT, we first broaden our descriptive analysis of alphabetized co-authorships over time using a data set which is more comprehensive and diverse than that employed by LT. We utilize data from the economics subject category of the Web of Science (WoS) provided by Clavariate Analytics. Our data set includes papers of the document type article ranging from 1990 to 2013. Our analysis uses only articles published before 2014, since each article needs a citation window of at least three years to allow reliable impact measurements. We collected citations up to the end of 2016. We make two adjustments to the original data set inasmuch as we keep only those journals that are listed in 2013 in the WoS economics subject category. We therefore exclude journals that have been discontinued or re-classified. Furthermore, we exclude journals with less than 100 listed articles. This lower limit is necessary to achieve sufficient statistical power for the separate regressions for each journal. The final data set consists of 207,159 articles published in 307 journals. Of these articles, 125,559 have at least two authors (61%). Building upon LT, we define three categories of multi-authored papers (strata): (i) all multi-authored articles; (ii) articles with two authors and (iii) articles with more than two authors. The latter two categories together constitute the first.

The left panel of Figure 1 confirms the trend of increasing numbers of authors and a decline in single-authored papers in economics. This has also been documented by Nowell and Grijalva (2011), Rath and Wohlrabe (2016) as well as Kuld and O'Hagan (2018). Articles

with more than two authors have become particularly more prevalent recently.

The alphabetization rate for multi-authored papers is around 70% in our sample (see Table 2) which is similar to the value reported by Waltman (2012) for economics.[1] Table 2 also shows that the alphabetization rate for articles with two authors (roughly 80%) is higher than the overall figure. The right panel of Figure 1 shows the development of alphabetized co-authorships over time, stratified by number of co-authors. The gold standard of alphabetized author order seems to be on the decline, especially since 2005. This holds true more or less for all author number strata.

The overall decline in alphabetized co-authorships appears to be driven by two effects: First, the alphabetization rate for all strata is declining over time, not just for one specific stratum. Second, strata with a structurally lower rate of alphabetization (higher order co-authorships) seem to be gaining in importance. This can be deduced from the strong increase in average authors per article. Despite the declining trend in alphabetization rates, economics still ranks as one of those disciplines where alphabetical author orders are most widespread (Waltman 2012). This is in contrast to other disciplines such as theoretical physics (50%), political science (61%), or statistics (56%). Whereas Waltman (2012) puts the overall alphabetization rate of economics papers at around 72%, our analysis further shows that this value seems to be driven by articles with two authors. Further research will reveal whether this result holds true for other disciplines. The results of Waltman (2012) suggest that there may be other academic fields with similar average numbers of authors to economics and alphabetization shares that are much closer to our estimates of economics articles, but with more than two authors.

Another source of heterogeneity in alphabetization rates is revealed in Figure 2. This illustrates the distributions of average shares of articles with alphabetized co-authors by journal. Much like the historical gold standard of alphabetized co-authorship, this is at best riddled with exceptions. Nevertheless, the kernel density estimation of the distribution reveals a maximum scattered around 90%.

The example of some elite journals commonly referred to as the top 5 (Card and DellaV-

[1] See also Henriksen (2019) for a case study for Danish economists.

igna 2013) shows why the phenomenon of alphabetization may be viewed with too much confidence if one focuses only on economics. In these cases, the rates of alphabetic order are consistently higher than average: *Journal of Political Economy*: 83%, *American Economic Review*: 91%, *Econometrica*: 94%, *Quarterly Journal of Economics*: 94%, *Review of Economic Studies*: 96%. These patterns of co- authorship alphabetizations should be subject to cautious examination, especially if they prove to have a meaningful impact on the citation performance of the article. Nonetheless, the decreasing relative significance of alphabetized co-authorship over time could already be interpreted as evidence refuting the hypothesis that alphabetization potentially increases citations - why else should the apparent publication equilibrium move away from a promising strategy. Theoretical arguments as to why such an alphabetized equilibrium may exist, have been put forth by Engers, Gans, Grant, and King (1999). These will be put to an empirical test in the following.

Table 1: Alphabetical Order for articles in economics journals

	Number of authors		
	> 1	$= 2$	> 2
Total	125,559	79,280	46,279
Percent alphabetical	70.11	79.74	53.63
Percent non-alphabetical	29.89	20.26	46.37

3 Empirical approach and results

3.1 Empirical approach

In order to test whether alphabetization has a statistically significant effect on citations, we estimate the following model

$$cit_i = \alpha + \beta AB_i + \gamma_1 pages + \gamma_2 pages^2 + \delta_1 age + \delta_2 age^2 + \gamma authors_i + \chi_j + year_t \qquad (1)$$

where β is the coefficient of interest. The variable AB is a dummy variable which is 1 if the authorship of an article i is stated in an alphabetical order and 0 otherwise. We include

Figure 1: Development of numbers of authors and author shares (left panel) and alphabetization rates (right panel) over time

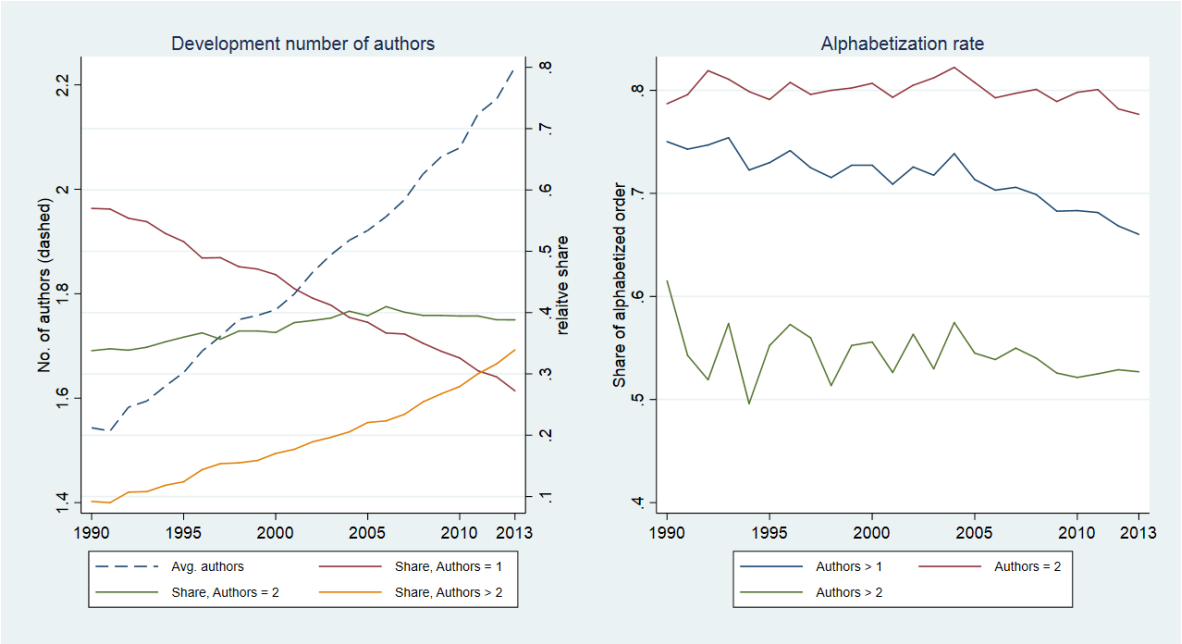
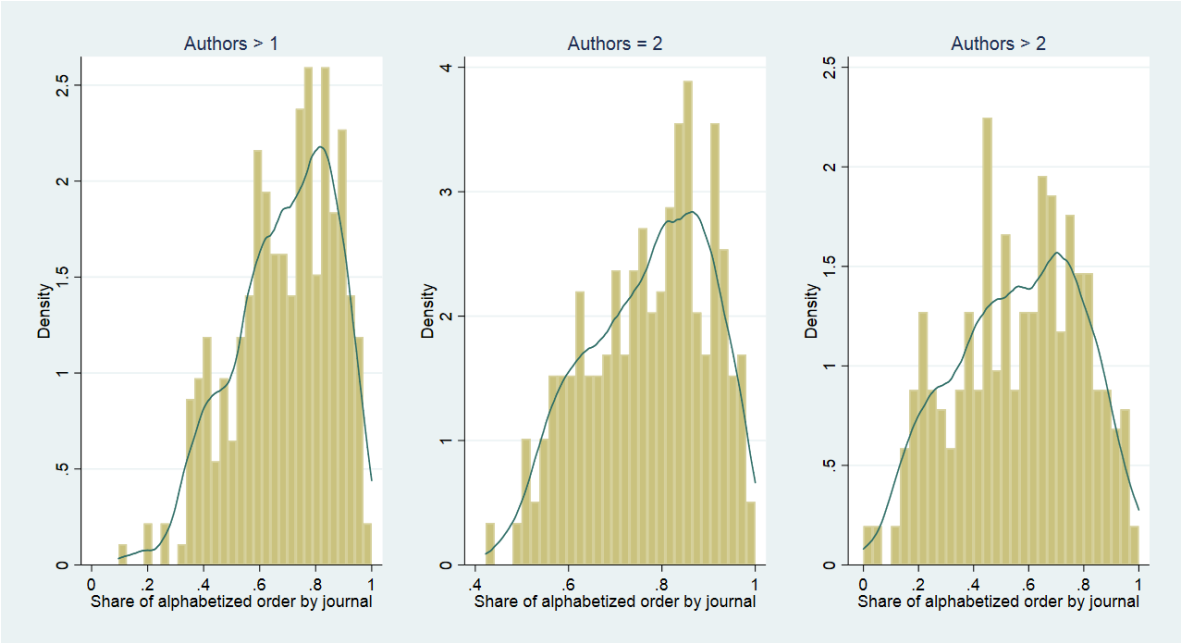


Figure 2: Histograms with kernel estimates of alphabetization rates by author strata across journals



further variables that might influence the number of citations up to the end of 2016: number of pages, article age (2016–publication year) and the number of authors. These variables have been frequently identified as factors that may influence citations in various studies (see the overview in Tahamtan and Bornmann 2018). The number of pages and article age are also included in the regression model as squared terms to capture non-linear effects. We also account for journal (χ_j) and time ($year_t$) fixed effects. The former accounts for specific journal quality and the latter for citation practices over time. We estimate four different models:

1. Negative binomial regression (NBR) model, where citations are interpreted as counts
2. basic OLS regression model
3. OLS using the natural log of citations as the dependent variable (this accounts for skewness in the citation distribution)
4. OLS using the inverse hyperpolic sine (IHS or $asinh$) transformation of citations, similar to log transformation, as proposed by Burbidge, Magee, and Robb (1988) and put forward recently by Card and DellaVigna (2017) in a citation analysis. The formal definition is $asinh(z) = \ln(z + \sqrt{1 + z^2})$. For $z \geq 2$, $asinh(z) = \ln(z) + \ln(2)$, but $asinh(0) = 0$.

The four approaches should yield robust results with respect to both the estimation approach and the handling of the dependent variable, the citations.

The results of the regression analyses which we present in the following are able to show whether there is a relationship between citation counts and alphabetical ordering of co-authors. It is not possible to reveal causal relationships between both variables. An important limitation of our design is that we do not (cannot) control the quality of papers (e.g. by including assessments of the papers by peers). Thus, we cannot completely exclude the possibility that differences in co-authorship ordering reflect systematically different quality levels.

3.2 Results

The results of the regression analysis are shown in Table 2 and are structured as follows. Each panel of Table 2 corresponds to an author number stratification. Each column within a panel then corresponds to one of four specifications of our regression model. In agreement with with previous literature (e.g. Gnewuch and Wohlrabe 2017), we obtain the expected sign of the coefficients for the explanatory variables and these are statistically significant in almost all cases. For instance, the longer an article or the more authors, the more citations an articles receives. With respect to our variable of interest we find no statistically significant effect of alphabetization on citations. This generally holds true for multiple authors, two authors and more than two authors. When interpreting the results, it should be considered that the sample size is very high, which makes statistically insignificant results unlikely. Thus, our findings are in contrast to the results reported in LT.

As the period of time investigated in LT is different and the degree of alphabetization has changed over time (see above), we additionally run the regression analysis for each year separately. The results for each author strata and specification are presented in Figure 3. This shows the estimated alphabetization coefficient (β) plus the 95% confidence interval bands. If the bands include the zero line, the coefficient is not statistically significant. The apparent statistical insignificance of the coefficient for alphabetized co-authorship for most of the years in our sample confirms the results in Table 2. However, there are some years in the 1990s with statistically significant coefficients that also correlate in part with the time frame in LT. However, the estimated significant coefficients are small. For example, in 1993, the OLS coefficient is around 0.25.[2] Thus, the effect of alphabetized co-authorship on citations is only marginal, even though it is statistically significant in some cases.

As a final check we repeat the analysis for each journal separately as was done for two journals in LT. Table 3 shows the relative shares of journals where the alphabetization dummy was statistically significant at different levels across different estimation approaches and degrees of co-authorship. There are journals with statistically significant effects of alphabetization on

[2]The corresponding coefficient in LT for the *American Economic Review* is about 32.

citations. In all cases the relative shares are somewhat higher than what can be expected from statistical theory.[3] Following LT, we take a closer look at the *American Economic Review* and the *American Journal of Agricultural Economics*. In the former case, we can confirm the results of LT. We find a statistically significant coefficient for the NBR and standard OLS with citations as the dependent variable for articles with two authors.[4] In all other specifications and author strata we find no statistically significant effects of alphabetization. For the *American Journal of Agricultural Economics*, the results do not point to statistically significant effects across all specifications. Thus, in this case we cannot confirm the results by LT.

In a final step we repeat the analysis of Brown, Chan, and Chen (2011) and ask whether the number of authors affects the probability of alphabetical ordering. We estimate a linear probability model using OLS and a logit regression by including the number of authors and pages as explanatory variables. Table 4 shows the corresponding estimation results. In line with Brown, Chan, and Chen (2011) we find that more authors increase the likelihood of authors being ordered *non*-alphabetically.

4 Discussion

There is no doubt as to the importance of citations. Citations promise to be an objective, long-term measure of a researcher's impact on the scientific discourse channeled by academic journals, as well as a proxy for the quality of precisely these journals. The important role of citations for both authors and journals naturally raises the question of what determines the number of citations per article? The quality of research contributions is an appealing answer. The literature, however, finds various other potential factors of a contextual (author or journal quality) as well as a technical nature, such as title characteristics (Gnewuch and Wohlrabe 2017), and subject category (Medoff 2003). Bornmann and Daniel (2008) provide an overview of citing behavior and factors that potentially influence citations. Laband and

[3] Based on our sample of 307 journals and given the three significance level 1%, 5% and 10%, one can expect approximately 3, 15 and 30 statistically significant results for the alphabetization parameter respectively.

[4] Please note that LT have partially different covariates in their regression models.

Table 2: Regression results stratified by number of authors

	(1)	(2)	(3)	(4)
Estimation approach	NBR	OLS	OLS	OLS
Dependent variable	Citations	Citations	Log citations	IHS citations
Authors > 1				
Alphabetical order of authors ($1 = yes$)	0.0013	0.0800	0.0001	0.0009
	(0.0103)	(0.3060)	(0.0068)	(0.0078)
Number of authors	0.0903***	1.807***	0.0788***	0.0898***
	(0.0047)	(0.1720)	(0.0034)	(0.0039)
Number of pages	0.0478***	0.9670***	0.0416***	0.0477***
	(0.0015)	(0.0488)	(0.0018)	(0.0021)
Number of pages (squared)	-0.0004***	-0.0041***	-0.0003***	-0.0004***
	(0.0000)	(0.0011)	(0.0000)	(0.0000)
Age of article	0.4450***	2.2900***	0.3180***	0.3830***
	(0.0231)	(0.2930)	(0.0140)	(0.0170)
Age of article (squared)	-0.0130***	-0.0388***	-0.0096***	-0.0116***
	(0.0008)	(0.0108)	(0.0005)	(0.0006)
Constant	0.0350	-3.6620	0.1770**	0.3200***
	(0.0990)	(3.8470)	(0.0576)	(0.0677)
Time and journal fixed effects	✓	✓	✓	✓
N	125535	125535	125535	125535
Authors $= 2$				
Alphabetical order of authors ($1 = yes$)	0.0075	0.0586	0.0059	0.0085
	(0.0133)	(0.3960)	(0.0090)	(0.0105)
Pages of article	0.0506***	0.9530***	0.0431***	0.0496***
	(0.0021)	(0.0602)	(0.0031)	(0.0037)
Number of pages (squared)	-0.0004***	-0.0042**	-0.0003***	-0.0004***
	(0.0000)	(0.0016)	(0.0000655)	(0.0001)
Age of article	0.4700***	1.8600***	0.3050***	0.3690***
	(0.0335)	(0.4000)	(0.0191)	(0.0232)
Age of article (squared)	-0.0138***	-0.0259	-0.00910***	-0.0111***
	(0.0011)	(0.0145)	(0.0007)	(0.0008)
Constant	-0.0016	-2.0450	0.2950***	0.4400***
	(0.1270)	(4.0360)	(0.0777)	(0.0920)
Time and journal fixed effects	✓	✓	✓	✓
N	79263	79263	79263	79263
Authors > 2				
Alphabetical order of authors ($1 = yes$)	-0.0002	0.2940	0.0006	0.0009
	(0.0150)	(0.4950)	(0.0104)	(0.0121)
Number of authors	0.0709***	1.477***	0.0604***	0.0674***
	(0.0062)	(0.2200)	(0.0048)	(0.0054)
Number of pages	0.0443***	1.0160***	0.0394***	0.0447***
	(0.0019)	(0.0843)	(0.0018)	(0.0021)
Number of pages (squared)	-0.0003***	-0.0041*	-0.0003***	-0.0003***
	(0.0000)	(0.0018)	(0.0000)	(0.0000)
Age of article	0.4270***	2.8080***	0.3390***	0.4060***
	(0.0302)	(0.4400)	(0.0207)	(0.0249)
Age of article (squared)	-0.0126***	-0.0549**	-0.0103***	-0.0124***
	(0.0010)	(0.0169)	(0.0007)	(0.0008)
Constant	0.3650*	2.4670	0.3070***	0.4880***
	(0.1560)	(8.1570)	(0.0859)	(0.0998)
Time and journal fixed effects	✓	✓	✓	✓
N	46272	46272	46272	46272

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3: Alphabetization impact over time across regression approaches

Note: Each graphs plots the estimated alphabetization parameter β plus the 95% confidence interval bands.

Table 3: Effect of alphabetization on citations across journals

Estimation approach	Dependent variable	Number of authors > 1	$= 2$	> 2
p-value=0.01				
NBR	Citations	3.26	4.89	5.54
OLS	Citations	1.30	3.26	1.30
OLS	Log citations	1.30	0.98	1.63
OLS	IHS citations	1.30	0.98	1.63
p-value=0.05				
NBR	Citations	10.10	12.70	13.03
OLS	Citations	6.19	8.14	5.86
OLS	Log citations	6.84	5.54	7.49
OLS	IHS citations	7.17	5.86	7.49
p-value=0.10				
NBR	Citations	17.92	18.24	19.87
OLS	Citations	13.68	13.36	11.73
OLS	Log citations	13.36	12.38	13.03
OLS	IHS citations	13.03	11.07	13.36

Notes: The table shows the relative shares of significant cases of the alphabetization parameter in regression (1) run over each journal for three significant levels.

Table 4: Regression results: determinants of alphabetization

	(1) OLS	(2) Logit
Number of pages	0.0003613	0.0035777
	(0.0005082)	(0.0029066)
Number of pages (squared)	0.0000121	0.0000805
	(0.00000984)	(0.0000578)
Number of authors	-0.1106878***	-0.8186007***
	(0.0027538)	(0.010939)
Time and journal fixed effects	✓	✓
N	125,535	125,481

Notes: Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tollison (2006) find that with up to two authors - alphabetization also pays off potentially in the form of more citations. Based on their analysis, the authors conclude that a team size of two is optimal. This is in line with the empirical fact that sole-authored articles in economics are in decline whereas multi-authored ones have become more and more popular. This result has been documented, for instance, by Nowell and Grijalva (2011) and more recently by Rath and Wohlrabe (2016) as well as Kuld and O'Hagan (2018). Huang (2015) estimates that economic papers whose authors' surname initials appear earlier in the alphabet receive more citations, thus suggesting a new dimension that has to be considered when choosing the order of authors. The finding that authors listed earlier in the alphabet may be favored in terms of academic rewards has been confirmed by several other authors: Einav and Yariv (2006), Efthyvoulou (2008) or Maciejovsky, Budescu, and Ariely (2009). Weber (2018) provides a survey of empirical evidence with respect to alphabetical ordering.

This article re-investigates the issues raised by LT using a much larger data set consisting of more than 120,000 multi-authored articles published in 307 journals from the economics category in WoS. In the first part of the study we show that the alphabetization rate in economics has declined since the early 2000s. This is in contrast with the prediction of Engers, Gans, Grant, and King (1999) who postulated a theoretical equilibrium of all authors playing an "alphabetized order strategy". In the second part we use four different regression settings to answer the question whether the order of authors affect the number of citations of an article. We do not find any statistically significant effects of alphabetization across all settings. Our findings thus directly contradict the conclusions in LT. However, our data set does confirm the finding of LT for the *American Economic Review* namely that alphabetical order articles with two authors obtain significantly more citations. This relationship also holds true for some other journals, though we believe this is more an artifact of the data. Our analysis shows that in the light of rising co-authorship (in economics), the ordering of authors does not matter, at least not for citations. Ray and Robson (2018) proposed an algorithm for random co-authorship listing. If this is adopted by all journals, an analysis like ours should become obsolete in the future.

Acknowledgments

We would like to thank Nikolai Schaffner for assisting the research presented in this paper.

References

Bornmann, L., A. Butz, and K. Wohlrabe (2018): "What are the top five journals in economics? A new meta-ranking," *Applied Economics*, 50(6), 659–675.

Bornmann, L., and H.-D. Daniel (2008): "What do citation counts measure? A review of studies on citing behavior," *Journal of documentation*, 64(1), 45–80.

Brown, C. L., K. C. Chan, and C. R. Chen (2011): "First-author conditions: Evidence from finance journal coauthorship," *Applied Economics*, 43(25), 3687–3697.

Burbidge, J. B., L. Magee, and A. L. Robb (1988): "Alternative transformations to handle extreme values of the dependent variable," *Journal of the American Statistical Association*, 83(401), 123–127.

Card, D., and S. DellaVigna (2013): "Nine Facts about Top Journals in Economics," *Journal of Economic Literature*, 51(1), 144–61.

——— (2017): "What do editors maximize? Evidence from four leading economics journals," NBER Working Papers 23282, National Bureau of Economic Research, Inc.

Efthyvoulou, G. (2008): "Alphabet economics: The link between names and reputation," *The Journal of Socio-Economics*, 37(3), 1266–1285.

Einav, L., and L. Yariv (2006): "What's in a surname? The effects of surname initials on academic success," *Journal of Economic Perspectives*, 20(1), 175–187.

Engers, M., J. S. Gans, S. Grant, and S. P. King (1999): "First-author conditions," *Journal of Political Economy*, 107(4), 859–883.

Gnewuch, M., and K. Wohlrabe (2017): "Title characteristics and citations in economics," *Scientometrics*, 110(3), 1573–1578.

Henriksen, D. (2019): "Alphabetic or contributor author order. What is the norm in Danish economics and political science and why?," *Journal of the Association for Information Science and Technology*, 70(6), 607–618.

Huang, W. (2015): "Do ABCs get more citations than XYZs?," *Economic Inquiry*, 53(1), 773–789.

Kuld, L., and J. O'Hagan (2018): "Rise of multi-authored papers in economics: Demise of the 'lone star' and why?," *Scientometrics*, 114(3), 1207–1225.

Laband, D., and R. Tollison (2006): "Alphabetized coauthorship," *Applied Economics*, 38(14), 1649–1653.

Maciejovsky, B., D. V. Budescu, and D. Ariely (2009): "The researcher as a consumer of scientific publications: How do name-ordering conventions affect inferences about contribution credits?," *Marketing Science*, 28(3), 589–598.

Medoff, M. H. (2003): "Article placement and market signalling," *Applied Economics Letters*, 10(8), 479–482.

Nowell, C., and T. Grijalva (2011): "Trends in co-authorship in economics since 1985," *Applied Economics*, 43(28), 4369–4375.

Rath, K., and K. Wohlrabe (2016): "Recent trends in co-authorship in economics: Evidence from RePEc," *Applied Economics Letters*, 23(12), 897–902.

Ray, D., and A. Robson (2018): "Certified Random: A new order for coauthorship," *American Economic Review*, 108(2), 489–520.

Tahamtan, I., and L. Bornmann (2018): "Core elements in the process of citing publications: Conceptual overview of the literature," *Journal of Informetrics*, 12(1), 203–216.

Waltman, L. (2012): "An empirical analysis of the use of alphabetical authorship in scientific publishing," *Journal of Informetrics*, 6(4), 700–711.

Weber, M. (2018): "The effects of listing authors in alphabetical order: A review of the empirical evidence," *Research Evaluation*, 27(3), 238–245.